



Towards a Discipline for Evaluating Ubiquitous Computing Applications

Jean Scholtz

NIST

&

Sunny Consolvo

Intel Research



Outline of Talk

- Motivation
- Ubicomp Evaluation Areas
 - Background
 - Discussion
- Examples
- Future work



Motivation

- Ubicomp evaluations are difficult
 - Hard to learn from previous studies
 - Evaluations are specifically designed for each application
 - Guidelines for design do not exist
 - Evaluations are conducted empirically
 - Expensive to construct field tests
 - Need robust software/hardware for field test
- A framework for evaluation would:
 - Make is easier to learn from each other
 - Enable creation of guidelines and “discount” methods of evaluation
 - Provide a way to share evaluation methodologies
 - Provide structure for planning evaluations



Related Work

- Jameson - 5 usability challenges for adaptive UIs
 - Predictability and transparency
 - Controllability
 - Unobtrusiveness
 - Privacy
 - Breath of experience
- We want to include adaptive systems but go beyond that
- Bellotti et al. - 5 interaction challenges for design of sensing systems
 - Address
 - Attention
 - Action
 - Alignment
 - Accident
- Our focus is on evaluator, not limited to GUI based interaction



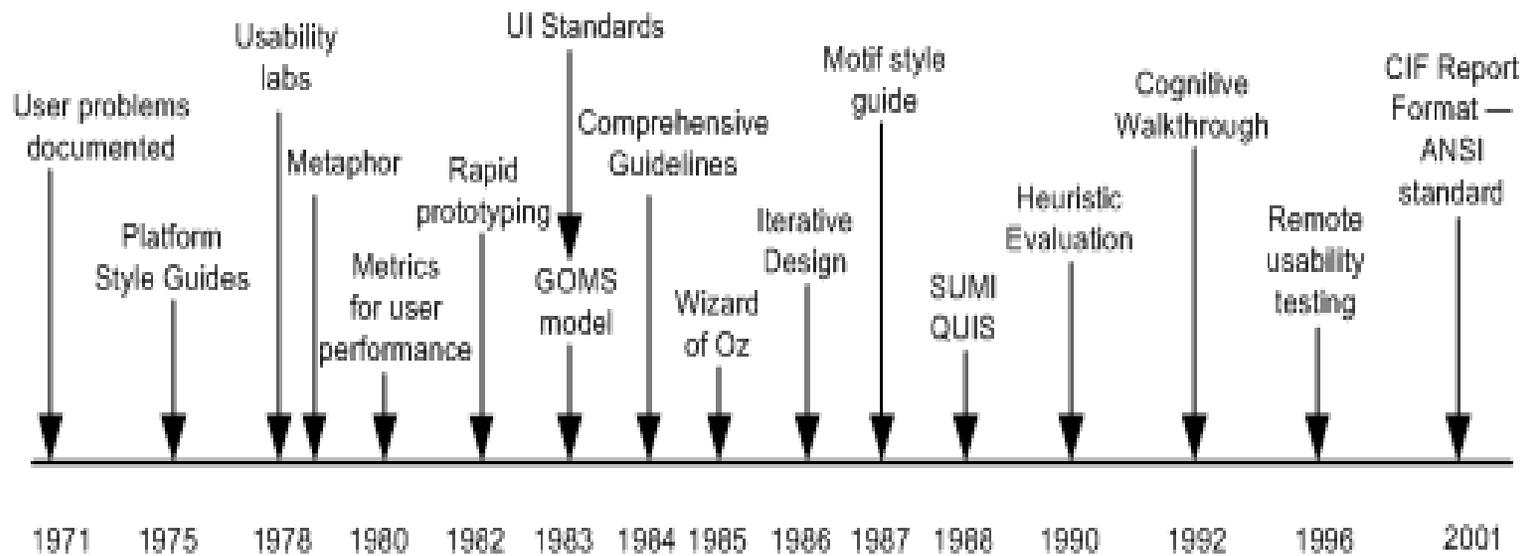
Traditional Desktop Usability

- Efficiency
 - Amount of time to perform a particular task
- Effectiveness
 - Percentage of task that the majority of users are able to complete with and without assistance
- User Satisfaction
 - How well users like using the system
- All of this is measured within the context of a task



Timeline 1971-2001

Information Access Division





What is Ubicomp?

- Ubicomp application is secondary to other tasks the user is performing
 - Shared with desktop computing but implies very different environments for computing
- Ubicomp environment may contain more than one interaction device that is used for any given application
- There are more varied interaction modalities
- A number of users may interact with the ubicomp application simultaneously
- Context-aware applications influence the behavior of an application



The Framework: Ubicomp Evaluation Areas (UEAs)

- Assembled from review of the literature and from personal experience
- Contain:
 - Definition
 - Metric: meaning associated with a measure
 - Conceptual measure: an observable value
 - Example(s)
- Metrics are used to compare two systems based on measures.
- Conceptual versus implementation measures
 - The actual measurement may differ depending on the application under consideration
- UEAs do overlap



Stakeholders

- Traditional usability evaluations focus on users
- In ubicomp applications we want to emphasize stakeholders
 - For example, Swarm- direct stakeholder is the person sending/receiving the message
 - Indirect stakeholder - people engaged in activities with the direct stakeholders while the interaction is occurring



UEA 1: Attention

- Attention: increased awareness directed at a particular event or action to select it for increased processing
- Metric: Focus
- Conceptual measures:
 - Number of times a user needs to change focus due to technology
 - Number of different displays/actions a user needs to accomplish an interaction or to check progress of an interaction
 - Number of events not noticed in an acceptable time



UEA 1: Attention, cont.

- Overhead
- Conceptual measures:
 - % time user spends switching foci
 - Workload imposed on the user due to changing foci



UEA 2: Adoption

- Should consider measuring the actual adoption but also predict success or failure
- Metric: Rate
- Conceptual measures:
 - New users/ unit of time
 - Adoption rationale
 - Technology usage statistics
- Metric: Value
- Conceptual measures:
 - Change in productivity
 - Perceived cost/benefit
 - Continuity for user
 - Amount of customer sacrifice



UEA 2: Adoption, cont.

- Metric: Availability
- Conceptual measures:
 - Number of actual users from each target group
 - Technology supply source
 - Categories of users in post-deployment



UEA 3: Trust

- Metric: Privacy
- Conceptual measures:
 - Amount of information user has to divulge to obtain value from application
 - Availability of explanations to user about use of recorded data
- Metric: Awareness
- Conceptual measures:
 - Ease of coordination with others in multi-user application
 - Number of collisions with activities of others
- Users often supply personal information to systems to make them more useful - this increases awareness but reduces privacy



UEA 4: Conceptual Models

- Metric: Predictability of application behavior
- Conceptual measures:
 - Degree of match between user's model and actual behavior
- Metric: Awareness of application capabilities
- Conceptual measures:
 - Degree of match between user's model and actual functionality



UEA 4: Mental Models, cont.

- Metric: Vocabulary awareness
- Conceptual measures:
 - Degree of match between user's model and syntax of interactions
 - What do I say to a smart room?



UEA 5: Interaction

- Metrics from traditional desktop computing are still needed
 - Effectiveness; efficiency; user satisfaction
- Metric: Distraction
- Conceptual measures:
 - Time taken from primary task
 - Degradation of performance in primary task
 - Level of user frustration
- Metric: Interaction transparency
- Conceptual measures:
 - Effectiveness comparison on different sets of input/output devices



UEA 5: Interaction, cont.

- Metric: Collaborative interaction
- Conceptual measures:
 - Number of conflicts
 - Percentage of conflicts resolved by the application
 - User feelings about the conflicts and resolutions
 - User ability to recover from conflicts



UEA 6: Invisibility

- Metric: Intelligibility
- Conceptual measures:
 - User's understanding of the system explanation
- Metric: Control
- Conceptual measures:
 - Effectiveness of interactions provided for user control of system initiative
- Metric: Accuracy
- Conceptual measures:
 - Match between system's contextual model and actual situation



UEA 6: Intelligibility, cont.

- Metric: Appropriateness of action
- Conceptual measures:
 - Match between system action and the action user would have requested
- Metric: Customization
- Conceptual measures:
 - Time to explicitly enter personalization information or time for system to learn and adapt



UEA 7: Impact

- Metric: Behavior changes
- Conceptual measures:
 - Type, frequency, and duration
 - Match between user's current job description and application role
- Metric: Social acceptance
- Conceptual measures:
 - Requirements placed on user outside of social norms
- Metric: Environment change
- Conceptual measures:
 - Type, frequency, and duration



Interpretation

- UEAs can be evaluated independently
- But we also need to look at combinations
 - Example: for context-aware applications we might get low control - if we also get low predictability of application behavior, the application may have problems with user acceptance.
 - Example: time critical applications need to score well in interaction and in attention.
 - Example: low scores on predictability of behavior; appropriateness of actions should predict that social acceptance will be low.
 - Example: large changes in behavior and social acceptance should correlate with adoption measures



Examples

- Evaluations published in Ubicomp 2002-2003
- Looked at what they did
 - Put their evaluation measures into our framework
 - Noted what fit and what didn't fit
 - Very informal - had to rely on what we could glean from paper



Burrell, 2002 - Campus Aware

- Tour guide for prospective students visiting campus
- Information from current campus community is used to populate
- Evaluation focused on "distraction"
 - Defined here as taking away from user's primary tasks - users found themselves watching the pda and not looking at the building
- Context-aware evaluation
 - Correlation between appearance of notes and physical objects
- UEA framework
 - UEA 1 : Attention - focus
 - UEA 5: Interaction - distraction
 - UEA 5: Intelligibility - accuracy



Burrell, 2002 - Campus Aware

- Scalability was also investigated
 - How to handle lots of comments, notes put in by others
- Our framework:
 - Does not address this
- We also suggest that evaluation should consider UEA 2: Adoption
 - Perceived cost/ benefit
 - This should be measured on community contributing comments as well
- And UEA 7: Impact
 - Social acceptance



Trevor, 2002 - Issues in Personalizing Shared Devices

- Evaluation considered
 - Trust and privacy
 - Availability of devices
 - Utility
- Our Framework:
 - UEA 3: Trust - privacy is considered one metric in this area
 - UEA 2: Adoption - availability measured by technology supply source
 - UEA 7: Impact - this would give utility along with UEA 2: Adoption - perceived benefit
 - UEA 5: Interaction - customization



Consolvo, 2002 - Labscape

- Study was based on interviews and contextual field research
- Identified work patterns that they used as design requirements
- Our framework:
 - UEA 7: Impact - behavior changes
 - UEA 1: Attention - focus



Darrell, 2002 - Face Responsive Interfaces

- Compared three prototypes:
 - Push to talk
 - Look to talk
 - Talk to talk
- Our framework:
 - UEA 5: Interaction - distraction and interaction transparency



Russell, 2002 - BlueBoard

- Field study of social effects of shared interactive displays
- Identified
 - Etiquette, who drives, learning by seeing, evolution of turn-taking, group sharing of information
- Our Framework:
 - UEA 4: Conceptual models - awareness of application capabilities, vocabulary
 - UEA 5: Interaction - collaborative interaction



La Marca, 2002 - PlantCare

- Evaluation was based on performance measures but says that impact on users needs to be measured
- Our Framework:
 - UEA 2: Adoption - perceived cost/benefit
Would address how much users will use to setup system based on perceived value



Barkhuus and Dey, 2003 - Control in Context-Aware

- Study looked at preferences for personalization, active context-aware behavior, and passive context-aware behavior in 7 different applications
- Found that personalization was used least; active context-aware was most preferred - but preference depended on type of service
- Our framework:
 - UEA 6: Invisibility - control
 - UEA 1: Attention - focus
 - There was an interaction here - users gave up control for benefit of having something automatically done
 - UEA 5: Interaction - transparency, customization



Salvador, 2003 - Practical Considerations of Context-aware

- This study is looking at designing for people with memory impairments
- Points out that user doesn't know the context - should system be allowed to infer it?
- Our Framework:
 - UEA 6: Control, appropriateness of action
 - But need to consider various categories of users



Cox, 2003 - IntelliBadge™

- Provided conference attendees with badges that allowed them to enter information and let others track them
- Provided lots of statistics about use
- Survey questions revealed
 - Need for more precise markers
 - Some did not see benefit
- Our Framework:
 - UEA 3: trust,
 - UEA 5: interaction - collaborative interaction
 - UEA 6: Intelligibility - customization
 - UEA 2: adoption - perceived cost/benefit
 - UEA 7: impact - social acceptance, behavior changes



Swarm

- Our framework:
 - UEA 5 - interaction
 - UEA 7 - impact, behavior changes, social acceptance
 - UEA 2 - adoption, cost/benefit, availability



What Next?

- We are missing UEAs for fun, enjoyment
- Metrics are focus on direct stakeholders - need to consider metrics for indirect stakeholders
- We need to do a thorough literature study and put some large studies into our framework
 - Formal version of my examples
- We need to conduct evaluations ourselves using these frameworks
 - UEA 4 - Conceptual models completed 2 studies trying to assess this
- We need also to look at evaluation methodologies for the different areas
 - Example: lag sequential analysis
- We would appreciate any help
 - Using framework
 - Providing us data to put into framework